DOCUMENT RESUME

ED 124 594 TM 005 353

AUTHOR Baker, Eva L.
TITLE Lean Data Strategies for Formative Evaluation.
PUB DATE [Apr 76]
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Data Collection; *Evaluation Methods; *Formative Evaluation; Instructional Improvement

ABSTRACT

The intent of formative evaluation is to improve programs as well as to justify their continuation. It is critical to separate clearly those functions of the evaluation which are political from those which may lead the way to instructional improvement. Data for formative evaluation should be gathered in an interpretable way at the level at which decisions will be implemented, usually at the classroom level. This forces data collection activities to a single planned incursion. This recommendation may suggest rethinking some of the measurement differences usually assumed to distinguish good classroom from good program evaluation. Purposive evaluation mandates asking questions about those areas over which a program developer or implementer exerts some control. Sometimes the specific implications of alternative data patterns are not kept in mind during the design of evaluations; the evaluator should attempt to foresee plausible alternative patterns the data might take. Rather than literally apply the evaluation plans of others, evaluation designs should be allowed to conform to the specific questions needing answers. What is desired are evaluation activities that have instructional improvement as well as, where necessary, political utilities; where with limited waste, data are provided that are pertinent and obvious to program needs. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED124594

LEAN DATA STRATEGIES FOR FORMATIVE EVALUATION

Eva L. Baker

Center for the Study of Evaluation

University of California, Los Angeles

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at a Symposium at the

Annual Meeting,

American Educational Research Association,

San Francisco, California, April, 1976.

TM005 353

LEAN DATA STRATEGIES FOR FORMATIVE EVALUATION

Eva L. Baker
Center for the Study of Evaluation
University of California, Los Angeles

At the outset, I wish to apologize for the title and in it, for the use of the term "lean". My adaptation of this word for application as a formative evaluation precept derived from one of Sue Markle's rules for the preparation of good programmed instruction. Programs were supposed to be written "leanly", that is, in minimal versions, so that field tests could provide information about what to add rather than require the more difficult inference about what to delete. "Lean" now strikes me as a word more appropriate for utterance by other scholars in the Chicago, specifically, Oscar Meyer and Farmer John, and I permanently consign the use of the term to them and their products.

Second, the economic constraints implied in the symposium title words, tight-money, suggest that the attribute of parsimony in evaluation strategy is something forced upon us, like margarine or Swanson Frozen Dinners because we can't afford to do things better and more expensively. I would propose to take the position that limits on the scope of evaluation are important for reasons other than thrift or deprivation and that the activity of evaluation itself demands concision of process.

Brief History

Although the activities now associated with evaluation had sludged along for years in the educational world, they coalesced

to form a defined field of inquiry only relatively recently (and we are all familiar with legislative and bureaucratic pressures that nominally gave evaluation activities their boost into prominence). The emergence of evaluation as something approaching a "field of inquiry" can be dated little more than ten years ago. Around that time people began to name themselves "evaluators", seeking an identifiable affiliation in doing so. Graduate programs appeared in which evaluation was legitimized as an endeavor worthy of an advanced degree. Scholars created sets of papers that attempted to define parameters of knowledge appropriate for evaluation. These papers were shared in elitist, ceremonial rites periodically and also published to inform a growing constituency. Followers or advocates of alternative models sprung up. The Charisma Coefficient, an indicator of professional efficacy emerged, in which the strength of following was directly related to the personal magnetism of the model-maker. Thus, we were presented with a series of papers describing alternative models and their variations, such as CIPP, Countenance, Discrepancy, Goal-Free and on and on. One confusing note in this period that has persisted was the ambiguity regarding what these models were for. Were they presented to help us organize the way in which we thought about evaluation, or was their purpose to control our actions and guide specific ways in which the evaluation was conducted?

These models may be regarded, then, as competitors for the hearts and minds of the people, however unlikely that intent in their development. The better models seemed to be more comprehensive and thorough.

In fact, certain models apparently arrogated functions that encompassed all of educational activity. For example, evaluation was seen to be critical at the point where goals were articulated, programs were planned, programs were implemented and results obtained and clearly, not only was the idea of evaluation important at each point, but detailed procedures to help evaluators conduct their business were also developed. Again, the more comprehensive the procedures, the better. For instance, these days in order to conduct a credible needs assessment, it is desirable to sample widely among real and imagined constituencies, so that students, teachers, administrators, parents, community members...all are represented. The needs assessment would then seem to be comprehensive, almost independent of whether those sampled had a sensible reason for being included at all. Because "limits to growth" was not a popular idea during the profligate sixties, the values of thoroughness and comprehensiveness were incorporated into referent works on evaluation and remain intact today.

Evaluation endeavors have evolved into a new meta-model. We have now firmly developed Procedure-Referenced, Decision-Free evaluation. As long as procedures are carefully followed, and we circumspectly remember to inflict them on all kinds of participants, our evaluation seems to meet the state-of-the-art, and stimulate familiar laments about why nobody uses what we do.

Kurt Vonnegut, in a book with an unpronouncible title, discussed speech-giving in the same way. No one cares or remembers anything

you say. What was important was that you seem to give a speech.

One of the contributing explanations for the procedure orientation of our evaluation efforts relates to the increasing sense that evaluation has primarily a political rather than a program improvement purpose. Evaluations are conducted because they are mandated, because we must justify previous decisions, because we need to show cause for increased funding. In the needs assessment example, we query multiple constituencies partly because we are interested in their perceptions but sometimes because we don't want to be accused of forgetting anyone. We involve complex procedures in our evaluations because they make us seem to be more credible. While it is difficult to argue that political factors are unimportant, they seem to confound our approach to evaluation so that, if those factors were put aside, we might not know the information we need to make a rational set of decisions anyhow.

Purpose and Plasticity

These remarks are directed, of course, not to the givers of evaluation models, but to the users of them. And the focus of the rest of the paper shall be on remedies in the content of formative evaluation.

With formative evaluation, our intent is to improve programs as well as to justify their continuation. Program improvement must focus at the classroom or learner level. Thus, it is critical in formative evaluation to separate clearly those functions of the evaluation which are political from those which may lead the way to

instructional improvement.

Specifically, what would that mean? Suppose one were interested in evaluating and improving a statewide educational program. One could develop and implement measures that were interpretable at the school district level, if the district were the main administrator of program. Such information would be helpful for the state evaluators and interesting to the district people. However, if the district wished to make decisions related to the improvement of school performance, school level information would be necessary. Thus the district might develop measures and a design to provide data for each participating school. Should a principal at a given school wish information, he or she would need to collect data for each of the classrooms in operation. And finally, should a given teacher in a classroom in a school in a district in an enlightened state wish information, he or she would obtain it from the students. Ultimately, for a program to produce improvement in learning, that teacher must make some good decisions about specific children. Our mythology is such that we believe that information about how the children are learning should help in that process.

Formative evaluation questions at the classroom/learner level seem to be rarely asked outside of organized, institutionally-based research and development efforts. Instead, each level of management, state, district, or school, seems to perpetrate another sort of evaluation effort.

Recommendation 1: Incursive evaluation

In the name of economy of time and animosity, data for formative evaluation should be gathered in an interpretable way at the level at which decisions will be implemented, usually at the classroom level. We need information about how specific children are doing on important outcomes. Such a desire certainly does not preclude sampling of persons or items or inhibit the manner of data reduction or aggregation useful for review at the subsequent administrative levels. Rather its force is to reduce data collection activities to a single carefully planned incursion. While the decision purposes for the evaluation data might be different for different users, various requirements may be attended to by the manner of aggregation and reporting rather than, as in many cases, the conduct of separate measurement activities. This recommendation may suggest rethinking some of the measurement differences usually assumed to distinguish good classroom from good program evaluation.

Recommendation 2: Purposive evaluation

Purposive evaluation invokes a simple mandate: ask questions about those areas over which, as a program developer or implementer, you exert some control Because of the procedural or political orientation of much of what we do, sometimes the specific implications of alternative data patterns are not kept in mind during the design of evaluations. One of the best residuals of our research

endeavors is the alternative hypothesis end-in-view. In research, we try to anticipate alternative hypotheses which can sensibly explain observations. In evaluation, it seems that the evaluator should attempt to foresee plausible alternative patterns the data might take. If the evaluator cannot imagine the consequences of data configurations for program improvement, then he or she might well consider whether certain questions need to be asked at all. For instance, if teacher age is found to be negatively correlated with program performance, but neither withholding the program from older teachers nor rejuvenation is possible in a cost-effective basis, then age-information, and many other demographic facts need not be assembled.

The recommendations for incursive and purposive evaluation designs suggest a retreat from procedure-oriented evaluation. Rather than apply literally the evaluation plans of others, our own evaluation designs should be allowed to conform to the specific questions we must have answered. Plasticity implies flexibility, moldability and idiosyncracy. Our evaluations will not seem as credible, because they will be based on local requirements rather than on cosmological methods. They will have to be justified by utility rather than authority. This should not suggest that evaluations are procedure-free. They may adapt well routinized methods for deciding on questions of importance and for obtaining, analyzing, and reporting our data. But the purpose controls the procedures.

Aside from cost savings derived from limiting measurements to few occasions and for consequence-related questions, support for stringent evaluation strategy comes from, of all places, an interpretation of the Second Law of Thermodynamics. Barry Commoner, in a recent article in the New Yorker, discusses the energy crisis with an analysis of the First and Second Laws of Thermodynamics. The First Law would suggest that, given a set of procedures, we attempt to implement them most efficiently. Thus, we might employ matrix sampling, and answer given questions in a way that saves both time and money. The Second Law, the law of entropy, suggests that every move we make contributes to the eventual randomness in the Universe. We would do best then, to rethink our questions and initiate the fewest activities required to provide answers, thereby creating the least disorder.

What we wish to develop, then, is virtual evaluation: evaluation activities that have instructional improvement as well as, where necessary, political utilities; where in a conserving way with limited waste, we provide data that are so pertinent and obvious to program needs that the lament that no one cares about evaluation is forgotten from disuse.